

Skript

HISTORISCH TIJDSCHRIFT

Column: 'Kun je 200.000 teksten niet gewoon zelf lezen?' Hoe digital humanities het historisch onderzoek veranderen

Auteur: Annemieke Romein

Verschenen in *Skript Historisch Tijdschrift*, jaargang 47.2, 32-34.

© 2025 Stichting Skript Historisch Tijdschrift, Amsterdam

e-ISSN 3051-2506

Skript Historisch Tijdschrift is een wetenschappelijk blad dat viermaal per jaar verschijnt. De redactie, bestaande uit studenten en pas afgestudeerden, wil bijdragen aan actuele historische debatten, en biedt getalenteerde studenten de kans om hun werk aan een breder publiek te presenteren.

'KUN JE DIE 200.000 TEKSTEN NIET GEWOON ZELF LEZEN?'

HOE DIGITAL HUMANITIES HET HISTORISCH ONDERZOEK VERANDEREN

Annemieke Romein

Een paar weken geleden kreeg ik op een bijeenkomst een open en geïnteresseerde vraag: 'Wat brengt het je nu, dat je middels *Automatic Text Recognition* (ATR)¹ en Transkribus zoveel teksten leesbaar kunt maken?' Geen waardeoordeel, geen scepsis – gewoon een nieuwsgierige vraag naar de impact van *digital humanities* op onderzoek. Een wereld van verschil met de reactie van een beoordelaar van een ERC-aanvraag (European Research Council) een paar jaar geleden. Die vroeg zich af waarom ik überhaupt financiering nodig had om 200.000 teksten te ontsluiten en analyseren – ik kon ze toch ook gewoon zelf even lezen? Zijn impliciete boodschap: technologie is een overbodige luxe, een trucje. Historisch onderzoek doe je met je ogen, je hersenen en een flinke dosis uithoudingsvermogen.

Maar *digital humanities* (digitale geesteswetenschappen) is veel meer dan een handige truc. Het is een veld dat de methodologische en technologische vooruitgang binnen de geesteswetenschappen stimuleert en historisch onderzoek ingrijpend verandert. Met technieken zoals ATR, *Natural Language Processing* (NLP) en geautomatiseerde classificatie kunnen onderzoekers sneller en systematischer werken, zonder noodzakelijkerwijs het traditionele vakmanschap te verliezen.

De kracht van schaal en automatisering

Een van de grootste veranderingen die digital humanities met zich meebrengen, is de mogelijkheid om onderzoek te doen op een veel grotere schaal zonder dat een onderzoeker al aan het eind van diens carrière hoeft te staan om het overzicht te hebben. Waar historisch onderzoek traditioneel vaak een diepgravende studie is

van een relatief beperkt aantal bronnen, kunnen we nu complete archiefcollecties analyseren en patronen ontdekken in datasets die voorheen onoverzichtelijk waren.

Denk bijvoorbeeld aan de Amsterdamse Akten², waarin dankzij ATR-technologie en *machine learning* een enorme hoeveelheid historische notariële documenten doorzoekbaar is gemaakt. Historici kunnen nu verbanden leggen tussen tienduizenden akten en zo bredere maatschappelijke ontwikkelingen in kaart brengen.

Ook binnen bibliotheken en archieven worden digital humanities steeds vaker ingezet om collecties beter toegankelijk te maken. Projecten zoals ANNIF (annif.org) laten zien hoe geautomatiseerde onderwerpherkenning kan helpen bij het toevoegen van metadata aan boeken, terwijl tools als OpenRefine (openrefine.org) onderzoekers helpen om

grote datasets te uniformeren en fouten in tabellen op te sporen. Dit maakt het mogelijk om sneller en nauwkeuriger te werken zonder dat men zich hoeft te verliezen in tijdrovende handmatige processen.

Digital humanities in de praktijk: Transkribus

Een concreet voorbeeld van hoe deze technologieën worden onderzocht, is het HAICu-project (haicu.science). Binnen dit onderzoeksproject wordt onderzocht hoe artificiële intelligentie (AI) gebruikt kan worden in de culturele sector. Zelf focus ik mij op het ontsluiten van de Resoluties van de Staten van Overijssel. Dit archief omvat zo'n 60.000 pagina's met vroegmoderne besluiten, vastgelegd in soms lastig leesbare handschriften en specifieke lay-outkeuzes. Dankzij ATR kan ik deze teksten omzetten naar doorzoekbare digitale bestanden. Dit bespaart niet alleen enorm veel tijd, maar maakt de bronnen ook toegankelijk voor een breder publiek: historici, genealogen, beleidsmakers en vooral de geïnteresseerde burger.

Tools zoals Transkribus (transkribus.eu), die gebruikmaken van ATR, laten zien hoe digital humanities niet alleen een kwantitatieve uitbreiding van onderzoek mogelijk maken, maar ook een kwalitatieve verbetering. Waar historici voorheen een steekproef namen uit archiefmateriaal, kan ik nu een volledig corpus systematisch analyseren. Dit betekent dat het niet alleen mogelijk is om te kunnen zoeken naar specifieke termen, ook het identificeren van bredere semantische structuren en het volgen van ontwikkelingen door de tijd heen wordt mogelijk.

Wat mij zelf enorm aanspreekt aan Transkribus is dat het een coöperatie is. Dit betekent dat iedereen aandelen kan kopen en als mede-eigenaar het platform in stand kan helpen houden. Niet alleen is het mogelijk om mee te beslissen, maar op deze manier wordt voorkomen dat de 'Big Tech' de techniek in handen krijgen. Het lijkt misschien utopisch, maar geeft

kleine talen óók de mogelijkheid om hun historisch materiaal te ontsluiten. Grote techbedrijven zijn voornamelijk geïnteresseerd in winst maken; terwijl Transkribus/READ-COOP de tools maakt waarmee onderzoekers/geïnteresseerden zelf aan de slag kunnen en hun eigen historisch materiaal kunnen ontsluiten. Bovendien ben ik van de generatie die nog geen programmeren leerde. Hoewel ik het meer en meer onder de knie krijg, vind ik het heel fijn dat de meeste processen in Transkribus geautomatiseerd zijn en ik dankzij het 'drukken op de knop' geen programmeerfouten kan maken. En ja, daar betaal je een bescheiden bedrag voor als je Transkribus op grote schaal gebruikt, maar dat valt in het niet als je berekent hoeveel het gebruiken van een server, onderhoud en ontwikkeling en personeel kost wanneer je het zelf zou draaien. Dat wordt vaak vergeten!

De symbiose tussen technologie en vakmanschap

Digital humanities bieden nieuwe mogelijkheden, hetgeen echter niet betekent dat traditioneel historisch onderzoek verdwijnt. Integendeel, de beste resultaten worden geboekt wanneer digitale technieken en klassiek onderzoek hand in hand gaan. De interpretatie van de gegenereerde data, de selectie van relevante onderzoeksmodellen en de uiteindelijke historische analyse blijven mensenwerk. Hoewel veel activiteiten door de computer vereenvoudigd kunnen worden – zoals het maken van een ATR-model – is het nog steeds handig om basale paleografische kennis te hebben om voor specifieke periodes zelf transcripties te kunnen maken en daarmee de computer te trainen. Bovendien, om de *inhoud* te begrijpen is veel historische en contextuele kennis noodzakelijk, waar nog steeds een degelijke opleiding voor nodig is.

Wat wel verandert, is de manier waarop historici werken. Voorheen was onderzoek grotendeels beperkt tot wat

men binnen een redelijke termijn kon lezen en verwerken. Nu kunnen we patronen ontdekken in eeuwen aan documenten, verbanden leggen die anders verborgen zouden blijven, en hypothesen testen op een schaal die voorheen ondenkbaar was. Deze nieuwe methoden vereisen wel dat historici nieuwe vaardigheden ontwikkelen. Basale programmeerkennis, begrip van datamodellen en vertrouwdsheid met digitale geletterdheid (*digital literacy*) worden steeds belangrijker in het curriculum van geschiedenisstudenten. Organisaties als Programming Historian (programming-historian.org) bieden laagdrempelige tutorials om deze vaardigheden te ontwikkelen.

De kern van de zaak

Digital humanities maken historisch onderzoek niet per se beter of slechter – ze maken het anders. Ze openen nieuwe mogelijkheden en stellen ons in staat om grootschaliger en systematischer te werk te gaan. Het traditionele handwerk zal niet verdwijnen, maar wel ontstaan er méér manieren om onderzoek te doen.

Dus, als ik die ERC-beoordelaar ooit zou spreken, zou ik willen antwoorden: ‘Nee, ik ga die 200.000 teksten niet één voor één lezen. Niet omdat ik dat niet zou kunnen, maar omdat ik dankzij digital humanities andere onderzoeksvragen kan stellen – vragen die ik op een traditionele

manier simpelweg niet zou kunnen beantwoorden.’ Digital humanities stellen me in staat om de bredere patronen te zien – en visualiseren – zonder de diepte te verliezen. En dat is geen gimmick – het is een fundamentele evolutie in hoe we als historici ons vak kunnen beoefenen.

Door het combineren van menselijke expertise met computationele methoden kunnen we het beste van twee werelden verenigen: de kritische blik van de historicus en de systematische analyse van technologie. En *dat* is precies wat nodig is om de complexiteit van ons verleden beter te begrijpen.

Annemieke Romein is historica, gespecialiseerd in vroegmoderne regelgeving. Ze onderzoekt hoe wetgeving regionaal verschildt en gebruikt daarbij digitale tools om duizenden documenten doorzoekbaar en vergelijkbaar te maken. Momenteel werkt zij bij het HAICu-project (UTwente)

Eindnoten

- 1 ATR is een combinatie van zowel *Optical Character Recognition* (OCR) als *Handwritten Text Recognition* (HTR). Deze technieken converteren respectievelijk gedrukte tekst of handgeschreven tekst om naar machineleesbare tekst (de output is dan bijvoorbeeld ‘een doorzoekbaar, veranderbaar wordbestandje’).
- 2 Raadpleegbaar via <https://www.amsterdam.nl/stadsarchief/alleamsterdamseakten/>.